Journal of Multivariate Analysis 108 (2012) 1–14

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Quantiles for finite and infinite dimensional data

Ricardo Fraiman [a,b], Beatriz Pateiro-López [c,*]

[a] Departamento de Matemática, Universidad de San Andrés, Argentina

[b] CMAT, Universidad de la República, Uruguay

[c] Departamento de Estadística e Investigación Operativa, Universidad de Santiago de Compostela, Spain

ARTICLE INFO

Article history:
Received 27 September 2010
Available online 9 February 2012

AMS subject classifications:
primary 62G30
secondary 62G20

Keywords:
Quantiles for functional data
Principal quantile directions
Hilbert space
High dimensional multivariate data

ABSTRACT

A new projection-based definition of quantiles in a multivariate setting is proposed. This approach extends in a natural way to infinite-dimensional Hilbert spaces. The directional quantiles we define are shown to satisfy desirable properties of equivariance and, from an interpretation point of view, the resulting quantile contours provide valuable information when plotting them. Sample quantiles estimating the corresponding population quantiles are defined and consistency results are obtained. The new concept of principal quantile directions, closely related in some situations to principal component analysis, is found specially attractive for reducing the dimensionality and visualizing important features of functional data. Asymptotic properties of the empirical version of principal quantile directions are also obtained. Based on these ideas, a simple definition of robust principal components for finite and infinite-dimensional spaces is also proposed. The presented methodology is illustrated with examples throughout the paper.

1. Introduction

The fundamental one-dimension concept of quantile function of a probability distribution is a well known device going back to the foundations of probability theory. The quantile function is essentially defined as the inverse of a cumulative distribution function. More precisely, given a real valued random variable X with distribution P_X, the α-quantile $(0 < \alpha < 1)$ is defined as

$$Q_X(\alpha) =: Q(P_X, \alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \tag{1}$$

where F denotes the cumulative distribution function of X. One prominent quantile value is $Q_X(0.5)$, the median of P_X, whose major role in probability theory and in mathematical statistics is well known. Nevertheless, the median is far from being the only important application of quantile functions. Quantiles of univariate data are the basis of the definition of other descriptive statistics as well as a powerful tool in estimation. In spite of the fact that the generalization of the concept of quantile function to a multivariate setting is not straightforward (due to the lack of a natural order in the d-dimensional space) a huge literature has been devoted to this problem in the last years. Different methodological approaches have been proposed, from those based on the concept of data depth (see for instance [25] or [35]) to those based on the geometric configuration of multivariate data clouds; see [5]. We refer to the survey by Serfling [32] for a complete overview and a exhaustive comparison of the different methodologies. Our proposal in this work is based on a directional definition of quantiles, indexed by an order $\alpha \in (0, 1)$ and a direction u in the unit sphere. To consider univariate quantiles for projections is quite a natural idea. An important contribution in this sense has been made recently by Kong and Mizera [24]. For a

* Correspondence to: Facultad de Matemáticas, Rúa Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain.
E-mail addresses: rfraiman@udesa.edu.ar (R. Fraiman), beatriz.pateiro@usc.es (B. Pateiro-López).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.01.016

given α, they start considering directional quantiles by projecting the probability distribution onto the straight line defined by each vector on the unit sphere. When projections in all directions are considered simultaneously a kind of summary is proposed as the "directional quantile envelopes", which turn out to be level sets of the half-space (Tukey) depth. More precisely, [19] observe that the approach in [24] yields a surprising connection between the quantile and depth philosophies in the multivariate setting. The inner regions characterized by the hyperplanes running through the α-quantile values and orthogonal to the directions for which they are calculated coincide with Tukey's halfspace depth regions. Hallin et al. [19] also define multivariate quantiles as hyperplanes whose inner regions coincide with the halfspace depth ones. Their quantile hyperplanes, however, are defined as regression quantile hyperplanes obtained in the traditional [23] sense and satisfy all desirable properties of equivariance. Anyway, as we shall see, if we first center properly the distribution it is easy to see that the more intuitive definition of directional quantiles will attain the main equivariance properties that are adequate for a quantile function.

On the other hand, beyond the lack of a widely accepted definition of multivariate quantiles there is also an increasing need for quantile functions valid for infinite-dimensional data (a problem recently posed by Jim Ramsay) in connection with the increasing demand of statistical tools for functional data analysis (FDA) where the available data are functions $x = x(t)$ defined on some real interval (say $[0, 1]$) see e.g., [13,12,30] or [14] for general accounts on FDA.

The idea of statistical depth can be extended to functional observations, see [15,27] among others. In [28] a new definition of depth for functional observations that provides a center-outward ordering of the sample curves is proposed. It also worth mentioning some recent outlier detection procedures for functional data based on functional depth measures; see [9,10,21]. Some of the ideas in [21] can be exploited in order to define new ordering graphical techniques using our definition of directional quantiles.

To sum up, the goal of this work is to provide an intuitive definition of directional quantiles that allow us to describe the behavior of a probability distribution in finite and infinite-dimensional spaces and provide some insight into the potential usefulness of our method.

2. Quantiles in Hilbert spaces. Definition and properties

We will work on the setup of Hilbert spaces (the simplest infinite dimensional vector space structure). Before stating the definition of quantiles in a Hilbert space, it is convenient to introduce some notation. In the remainder of this paper, $\mathcal{X}$ will denote a functional random variable valued in some infinite-dimensional space $\mathcal{E}$. We do not bother to distinguish in our notation between functions, scalar quantities and non-random elements of $\mathcal{E}$ and we use standard letters for all cases. Since we will still need to introduce multivariate variables in some definitions and examples, we adopt the convention of writing vectors as boldface lower case letters and matrices in boldface upper case. Let $\mathcal{H}$ be a separable Hilbert space where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\|\cdot\|$ denotes the induced norm in $\mathcal{H}$. Let $\mathcal{X}$ be a random element in $\mathcal{H}$ with distribution $P_{\mathcal{X}}$ and such that $\mathbb{E}(\|\mathcal{X}\|) < \infty$. Our extension of the concept of quantiles to multidimensional and infinite-dimensional spaces is based on a directional definition of quantiles. Thus, we denote $\mathbb{B} = \{u \in \mathcal{H} : \ \|u\| = 1\}$ the unit sphere in $\mathcal{H}$ and define, for $0 < \alpha < 1$, the α-quantile in the direction of $u \in \mathbb{B}$, $Q_{\mathcal{X}}(\alpha, u) \in \mathcal{H}$, as

$$Q_{\mathcal{X}}(\alpha, u) = Q_{\langle \mathcal{X} - \mathbb{E}(\mathcal{X}), u\rangle}(\alpha) u + \mathbb{E}(\mathcal{X}). \tag{2}$$

In some sense, this definition reminds us the quantile's definition (in a finite-dimensional setting) given by Kong and Mizera [24]. They define directional quantiles as the quantiles of the projections of the probability distribution into the directions of the unit sphere. However note that, in (2), the α-quantile in the direction of $u \in \mathbb{B}$ is defined from the α-quantile of the corresponding projection of $\mathcal{Z} = \mathcal{X} - \mathbb{E}(\mathcal{X})$. Centering the random element before projecting is essential in order to obtain quantile functions fulfilling desirable equivariance properties. Now, let $P_{\mathcal{Z}}(u)$ denote the probability distribution of the random variable $\langle \mathcal{Z}, u\rangle$. Following the notation introduced in (1) for the univariate case, the α-quantile in (2) can also be written as

$$Q_{\mathcal{X}}(\alpha, u) = Q(P_{\mathcal{Z}}(u), \alpha) u + \mathbb{E}(\mathcal{X}). \tag{3}$$

For convenience, we will use both the notations (2) and (3) throughout this paper. For fixed α, the quantile function $Q_{\mathcal{X}}(\alpha, \cdot)$ indexed by u in the unit sphere naturally yields quantile contours $\{Q_{\mathcal{X}}(\alpha, u), \ u \in \mathbb{B}\}$. Representing the quantile contours for different values of α may be an informative way of describing the behavior of the underlying probability distribution.

2.1. Equivariance properties

When defining quantile functions, as happens with other location measures, we expect changes on the original variables such as translations or homogeneous scale transformations, to have no fundamental effect on such functions. In other words, when the variables are altered in one of these ways, we expect quantiles also to change in a way that leaves our interpretation of the results unchanged.

The quantiles in (2) fulfill the following equivariance properties: location equivariance, equivariance under unitary operators and equivariance under homogeneous scale transformations.

All agree that any statistical procedure should depend "only on the data" and not on the coordinate system in which they are provided. This requirement will be fulfilled if the procedure is equivariant with respect to translations

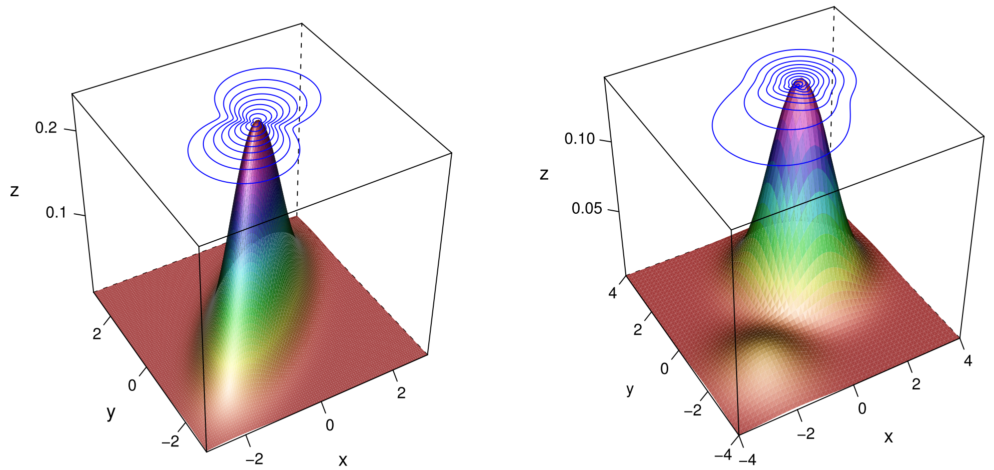

Fig. 1. Three-dimensional views of two distributions in $\mathbb{R}^2$ with two-dimensional quantile contours for $\alpha = 0.5, 0.55, \dots, 0.95$ projected onto the top. Left, Normal distribution with zero mean and covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})$, $\sigma_{ii} = 1$, $\sigma_{ij} = 0.75$, $i \neq j$. Right, mixture of two Normal distributions $N(\boldsymbol{\mu_i}, \boldsymbol{I})$ in $\mathbb{R}^2$. The mean vectors are $\boldsymbol{\mu}_1^t = (1, 1)$ and $\boldsymbol{\mu}_2^t = (-2, -2)$ and the mixing proportions are $p_1 = 0.9$ and $p_2 = 0.1$.

(does not depend on the origin of coordinates), unitary operators (does not depend on the orthogonal system of coordinates), homogeneous scale transformations (does not depend on the global scale of the data). The proposed quantiles satisfy all these desirable equivariance properties. Quite often an extra equivariance property in the finite-dimensional case is required: the affine equivariance property. Affine equivariance is natural in some situation like in the linear model, in order to take into account possible linear re-parametrizations of the model. However, this is not a mandatory property. In our setting the extra requirement of being affine equivariant is not clear. As an illustration, let us consider the case of an elliptical distribution, see for example Fig. 1 (left). There is a clear concentration of mass close to the origin in the direction $y = -x$. This relevant fact is also evident looking at the quantile contours derived from our definition of quantile function. On the contrary, this fact will not be shown properly by other multivariate quantile functions which are affine equivariant. Anyway, one could still define an affine equivariant version of the quantile function in the finite-dimensional case. This could be achieved through a standard normalization premultiplying the random vector by $\Sigma^{-1/2}$ and then premultiplying the quantile by $\Sigma^{1/2}$, where Σ is a (robust or not) scatter functional.

Based on our definition of quantiles, we propose in Section 4 the new concept of principal quantile directions, which is closely related in some situations with the principal component analysis. The results of PCA are not invariant under general affine transformations, in particular under changes in the units of the variables, since PCA is based on a fixed metric which is invariant only under orthogonal transformations. We refer to [29] for a deeper discussion on this topic.

2.2. Quantile contours in the multivariate setting

The preceding definition of quantiles in a separable Hilbert space applies directly to the euclidean space $\mathbb{R}^d$. Following the notation for the finite-dimensional case, if $\boldsymbol{X}$ is a random vector with $\mathbb{E}(\boldsymbol{X}) < \infty$, then

$$Q_{\boldsymbol{X}}(\alpha, \boldsymbol{u}) = Q_{\langle \boldsymbol{X}-\mathbb{E}(\boldsymbol{X}), \boldsymbol{u}\rangle}(\alpha)\boldsymbol{u} + \mathbb{E}(\boldsymbol{X})$$

denotes the α-quantile in the direction of $\boldsymbol{u} \in \mathbb{B}$. As in the general case, these directional quantiles yield quantile contours $\{Q_{\boldsymbol{X}}(\alpha, \boldsymbol{u}),\ \boldsymbol{u} \in \mathbb{B}\}$. Our objective here is to study the quantile contours of some relevant distributions. Although classical multivariate analysis has been built up on the assumption of normality of the observations, the fact is that real data very seldom satisfy this assumption. Spherically symmetric distributions and elliptically symmetric distributions are natural extensions to the multivariate standard normal and the general multivariate normal, respectively and have been used in many statistical models. We consider these two types of symmetry and study the quantile contours that result in both cases. The proofs of the facts stated are straightforward.

2.2.1. Spherical distributions

Let $\boldsymbol{X}$ be a random vector with finite expectation. The distribution of $\boldsymbol{X}$ is said to be spherically symmetric about $\mathbb{E}(\boldsymbol{X})$ if the distributions of $\boldsymbol{X} - \mathbb{E}(\boldsymbol{X})$ and $\boldsymbol{A}(\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}))$ are identical, for any orthogonal matrix $\boldsymbol{A}$. We state in Fact 1 below that, if a probability distribution is spherically symmetric about its expectation, $\mathbb{E}(\boldsymbol{X})$, then the α-quantile contour associated with that distribution will be a sphere with center in $\mathbb{E}(\boldsymbol{X})$ and radius that depends on α.

Fact 1. Let $\boldsymbol{X}$ be a random vector with finite expectation and spherically symmetric distribution about $\mathbb{E}(\boldsymbol{X})$ and let $0 < \alpha < 1$. Then, for all $\boldsymbol{u} \in \mathbb{B}$,

$$Q_{\boldsymbol{X}}(\alpha, \boldsymbol{u}) = c_\alpha \boldsymbol{u} + \mathbb{E}(\boldsymbol{X}).$$

Examples of spherically symmetric distributions are, among others, multivariate normal distributions with covariance matrices of form $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ and certain cases of standard multivariate t and logistic distributions. See [33] for more examples and useful characterizations of spherical symmetry.

2.2.2. Elliptical distributions

Other models that have received much attention are the elliptical distributions or elliptically contoured distributions. Let $\boldsymbol{X}$ be a random vector with finite expectation. The distribution of $\boldsymbol{X}$ is said to be elliptically symmetric if there exists a nonsingular matrix $\boldsymbol{B}$ such that $\boldsymbol{X} = \boldsymbol{BZ}$ where $\boldsymbol{Z}$ is spherically symmetric about $\mathbb{E}(\boldsymbol{Z})$. We have the following in view of the definition of quantiles.

Fact 2. Let $\boldsymbol{X}$ be a random vector with finite expectation and elliptically symmetric distribution. Let $0 < \alpha < 1$. Then, for all $\boldsymbol{u} \in \mathbb{B}$,

$$Q_{\boldsymbol{X}}(\alpha, \boldsymbol{u}) = \|\boldsymbol{B}^t \boldsymbol{u}\| c_\alpha \boldsymbol{u} + \mathbb{E}(\boldsymbol{X}).$$

From the definition, it is immediate that spherically symmetric distributions are also elliptically symmetric. As stated in Fact 1, the quantile contours given by Fact 2 for spherically symmetric distributions reduce to spheres with center in $\mathbb{E}(\boldsymbol{X})$. Note that, even though for elliptically symmetric distributions the contours of equal density are elliptical in shape, this does not occur, in general, with the quantile contours. As an example, see Fig. 1 (left) where the quantile contours of multivariate normal distributions in $\mathbb{R}^2$ are represented.

Finally and as an illustration, we show in Fig. 1 (right) the quantile contours of a mixture of Normal distributions in $\mathbb{R}^2$. Mixtures of Normal distributions enable us to generate a rich class of densities that accommodate modeling in situations where data exhibits multimodality and are useful in several practical applications. We calculate the quantile function by using that the linear projection onto $\boldsymbol{u}$ of the mixture of k Normal distributions $N(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ in $\mathbb{R}^d$, is also a mixture of k Normals with component mean and variance given by $\boldsymbol{u}^t \boldsymbol{\mu_i}$ and $\boldsymbol{u}^t \boldsymbol{\Sigma_i} \boldsymbol{u}$, $i = 1, \ldots, k$.

3. Sample quantiles

In order to define the sample version of the quantiles, let us first consider the univariate case. Given the observations $X_1, \ldots, X_n$, denote by P_n the empirical measure, that is, the random measure that puts equal mass at each of the n observations. For $0 < \alpha < 1$, the sample α-quantile, $Q(P_n, \alpha)$, is defined as

$$Q(P_n, \alpha) = \inf\{x \in \mathbb{R} : F_n(x) \geq \alpha\}, \tag{4}$$

where F_n denotes the sample cumulative distribution function,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}.$$

Clearly, if $X_1, X_2, \ldots, X_n$, are independent and identically distributed observations from a random variable X with distribution P_X, then $Q(P_n, \alpha)$ will act as an estimate of $Q_X(\alpha)$ based on those observations.

For the general setting, let $\mathcal{X}$ be a random element in $\mathcal{H}$ with probability distribution $P_{\mathcal{X}}$ such that $\mathbb{E}(\|\mathcal{X}\|) < \infty$. Then, let $\mathcal{Z} = \mathcal{X} - \mathbb{E}(\mathcal{X})$ with distribution $P_{\mathcal{Z}}$. Given $\mathcal{X}_1, \ldots, \mathcal{X}_n$ a random sample of elements identically distributed as $\mathcal{X}$, denote $\mathcal{Z}_{ni} = \mathcal{X}_i - \bar{\mathcal{X}}$, $i = 1, \ldots, n$. Now, for $u \in \mathbb{B}$, let $P_n(u)$ denote the empirical measure of the observations $\langle \mathcal{Z}_{n1}, u \rangle, \ldots, \langle \mathcal{Z}_{nn}, u \rangle$. We define the empirical version of the quantiles in (2) by replacing the univariate α-quantile, $Q_{\langle \mathcal{X} - \mathbb{E}(\mathcal{X}), u \rangle}(\alpha)$, with the sample α-quantile $Q(P_n(u), \alpha)$ as given in (4). That is, we define

$$\hat{Q}_{\mathcal{X}}(\alpha, u) = Q(P_n(u), \alpha) u + \bar{\mathcal{X}}, \tag{5}$$

where now

$$Q(P_n(u), \alpha) = \inf\{x \in \mathbb{R} : F_n^u(x) \geq \alpha\}$$

and

$$F_n^u(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{\langle \mathcal{Z}_{ni}, u \rangle \leq x\}}.$$

In practice, a direct translation of Eq. (5) involves computing infinitely many univariate quantiles. We propose to compute the empirical quantile function for a considerable number of directions in order to make the approximation satisfactory. This procedure is not troublesome if the dimension of the space which contains the data is not large. Otherwise, for instance if we work with functional data, a way to circumvent the computational difficulties is to employ randomly chosen projections. We refer to [6,7] for further exploration of the use of random projections for functions.

3.1. Asymptotic behavior

Before we tackle the asymptotic behavior of the sample quantiles $\hat{Q}_{\mathcal{X}}(\alpha, u)$ in (5), we will need some auxiliary results on the convergence of the empirical measure $P_n(u)$. Classical results on the consistency of the univariate sample quantiles are obtained as a consequence of the consistency of the empirical distribution function. However, the consistency of the empirical distribution function relies on the assumption of independent and identically distributed random variables, which is not the case in our setting. Note that in the definition of $Q(P_n(u), \alpha)$, the empirical distribution function is computed from the observations $\langle \mathcal{Z}_{n1}, u \rangle, \dots, \langle \mathcal{Z}_{nn}, u \rangle$, which are clearly not independent. For each $h \in \mathcal{H}$ denote by $F^h(t)$ the probability distribution function of the random variable $\langle \mathcal{Z}, h \rangle$. We obtain the following result, whose proof can be found in Section 7.

Proposition 1. *Let $\mathcal{H}$ be a separable Hilbert space. Then,*

$$\lim_{n\to\infty} \sup_{\|h\|=1, t\in\mathbb{R}} |F_n^h(t) - F^h(t)| = 0 \quad a.s. \tag{6}$$

if and only if

$$\lim_{\epsilon\to 0} \sup_{\|h\|=1, t\in\mathbb{R}} P(\{x \in \mathcal{H} : |\langle h, x\rangle - t| < \epsilon\}) = 0. \tag{7}$$

It can be proved that, for the Euclidean space $\mathbb{R}^d$, Condition (7) is straightforwardly satisfied. This result is presented in the following corollary.

Corollary 1. *For $\mathcal{H} = \mathbb{R}^d$, then*

$$\lim_{n\to\infty} \sup_{\|\boldsymbol{h}\|=1, t\in\mathbb{R}} |F_n^{\boldsymbol{h}}(t) - F^{\boldsymbol{h}}(t)| = 0 \quad a.s.$$

Remark 1. Proposition 1 gives us a necessary and sufficient condition for $\mathcal{F}$ to be a Pólya class for P, see the proof of this result in Section 7. This condition is fulfilled when $\mathcal{H} = \mathbb{R}^d$ as it is shown in Corollary 1. Unfortunately, Condition (7) does not hold for most interesting infinite-dimensional spaces. As an example, let $\mathcal{H} = L^2[0, 1]$ and $\mathcal{X}$ be a Gaussian process in $\mathcal{H}$, with $\mathbb{E}(\mathcal{X}) = 0$ and $E(\|\mathcal{X}\|^2) < \infty$. Let T be the (compact) covariance linear operator associated with $\mathcal{X}$. In this case we have that $\langle \mathcal{X}, h\rangle$ is a $N(0, \langle h, Th\rangle)$ random variable. On the other hand, we have that

$$\sup_{\|h\|=1, t\in\mathbb{R}} P\left(|\langle \mathcal{X}, h\rangle - t| < \epsilon\right) \geq \sup_{\|h\|=1} P\left(\frac{|\langle \mathcal{X}, h\rangle|}{\sqrt{\langle h, Th\rangle}} < \frac{\epsilon}{\sqrt{\langle h, Th\rangle}}\right) = 1,$$

since $\inf_{\|h\|=1} \langle h, Th\rangle = 0$, and condition (7) is not fulfilled. If we consider more restrictive infinite-dimensional spaces (for instance compact spaces) then (6) holds. Some conditions under which the result is valid are given in [1, Illustration 4.1].

The pointwise consistency of $Q(P_n(u), \alpha)$ to $Q(P_{\mathcal{Z}}(u), \alpha)$ (for each fixed direction u) is stated in Proposition 2. Proposition 3 establishes the uniform convergence of $Q(P_n(u), \alpha)$. Finally, the uniform convergence of the sample quantiles to the population version is obtained in Proposition 4.

Proposition 2. *Let $\mathcal{H}$ be a separable Hilbert space and $0 < \alpha < 1$. With the previous notation assume that,*

(i) *given $u \in \mathbb{B}$, $F^u(t) > \alpha$ for all $t > Q(P_{\mathcal{Z}}(u), \alpha)$,*
(ii) $\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |F_n^u(t) - F^u(t)| = 0$ *a.s.*

Then, for each $u \in \mathbb{B}$,

$$\lim_{n\to\infty} |Q(P_n(u), \alpha) - Q(P_{\mathcal{Z}}(u), \alpha)| = 0 \quad a.s. \tag{8}$$

Proposition 3. *Let $\mathcal{H}$ be a separable Hilbert space and $0 < \alpha < 1$. With the previous notation assume that,*

(i) *given $\epsilon > 0$, there exists $\delta_0 > 0$ such that*

$$\sup_{\|u\|=1} F^u(Q(P_{\mathcal{Z}}(u), \alpha) - \epsilon) < \alpha - \delta_0 \tag{9}$$

and

$$\sup_{\|u\|=1} F^u(Q(P_{\mathcal{Z}}(u), \alpha) + \epsilon) > \alpha + \delta_0. \tag{10}$$

(ii) $\lim_{n\to\infty} \sup_{\|u\|=1, t\in\mathbb{R}} |F_n^u(t) - F^u(t)| = 0$ *a.s.*

Then,

$$\lim_{n\to\infty} \sup_{\|u\|=1} |Q(P_n(u), \alpha) - Q(P_{\mathcal{Z}}(u), \alpha)| = 0 \quad a.s.$$

Proposition 4. *Under the conditions of Proposition 3,*

$$\lim_{n\to\infty} \sup_{\|u\|=1} \left\| \hat{Q}_{\mathcal{X}}(\alpha, u) - Q_{\mathcal{X}}(u, \alpha) \right\| = 0 \quad a.s.$$

4. Principal quantile directions

One of the goals of the multivariate data analysis is the reduction of dimensionality. The use of principal components is often suggested for such dimensionality reduction. Thus, Principal Component Analysis (PCA) is one of the most widely used multivariate techniques of exploratory data analysis. Comprehensive coverage of this topic and references can be found in the book by Jolliffe [22]. More recently, the PCA methods were extended to functional data and used for many different statistical purposes; see [30]. Research on functional PCA includes that of [4,16,17], among others. More recent work addresses the application of functional PCA to different fields, such as longitudinal data analysis; see [18].

A way to summarize the information in the quantile functions is to consider principal quantile directions for a given level α, defined as follows. The first principal quantile direction is the one that maximizes the norm of the centered quantile function $Q_{\mathcal{X}}(\alpha, u) - \mathbb{E}(\mathcal{X})$, i.e. the direction $u_1 \in \mathbb{B}$ satisfying

$$u_1 = \arg\max_{u\in\mathbb{B}} \left| Q_{\langle \mathcal{X}-\mathbb{E}(\mathcal{X}),u\rangle}(\alpha) \right|. \tag{11}$$

The second principal quantile direction is defined as in principal components. It is the direction $u_2 \in \mathbb{B}$ satisfying

$$u_2 = \arg\max_{u\in\mathbb{B}, u\perp u_1} \left| Q_{\langle \mathcal{X}-\mathbb{E}(\mathcal{X}),u\rangle}(\alpha) \right|.$$

The k-principal quantile direction is defined as the direction $u_k \in \mathbb{B}$ satisfying

$$u_k = \arg\max_{u\in\mathbb{B}, u\perp H_{k-1}} \left| Q_{\langle \mathcal{X}-\mathbb{E}(\mathcal{X}),u\rangle}(\alpha) \right|, \tag{12}$$

where H_{k-1} is the linear subspace generated by $u_1, \ldots, u_{k-1}$.

Since the unit ball is weakly compact (compact with respect to the weak topology), the maximum is attained whenever $\left| Q_{\langle \mathcal{X}-\mathbb{E}(\mathcal{X}),u\rangle}(\alpha) \right|$ is weakly continuous as a function of u, and principal quantile directions are well defined.

As mentioned before, in many situations the classical Principal Component Analysis is considered as a useful tool for displaying data in a reduced dimensional space. Next results show situations in which the principal quantile directions coincide with the principal components.

Proposition 5. *Let $\boldsymbol{X}$ be a random vector with finite expectation and elliptically symmetric distribution. Then, the principal quantile directions defined by (11) and (12) coincide with the principal components.*

Proposition 6. *Let $\mathcal{X} = \{X(t),\ t \in [0, 1]\}$ be a Gaussian process in $L^2[0, 1]$ with covariance function*

$$\gamma(s, t) = \mathrm{Cov}(X(t), X(s)),$$

which we assume to be square integrable. Then, the principal quantile directions defined by (11) and (12) coincide with the principal components. Moreover,

$$\max_{u\in\mathbb{B}, u\perp H_{k-1}} \left| Q_{\langle \mathcal{X}-\mathbb{E}(\mathcal{X}),u\rangle}(\alpha) \right| = \Phi^{-1}(\alpha)\sqrt{\lambda_k}, \tag{13}$$

where Φ stands for the cumulative distribution function of a standard Normal random variable and $\lambda_1 \geq \lambda_2, \ldots$ is the sequence of eigenvalues of the covariance operator $\Gamma : L^2[0, 1] \to L^2[0, 1]$ defined as

$$\Gamma(u)(s) = \int_0^1 u(t)\gamma(s, t)dt.$$

4.1. Sample principal quantile directions

The first sample principal quantile direction is the one that maximizes the norm of the centered empirical quantile function $Q(P_n(u), \alpha)$, i.e. the direction $\hat{u}_1 \in \mathbb{B}$ satisfying

$$\hat{u}_1 = \arg\max_{u\in\mathbb{B}} |Q(P_n(u), \alpha)|.$$

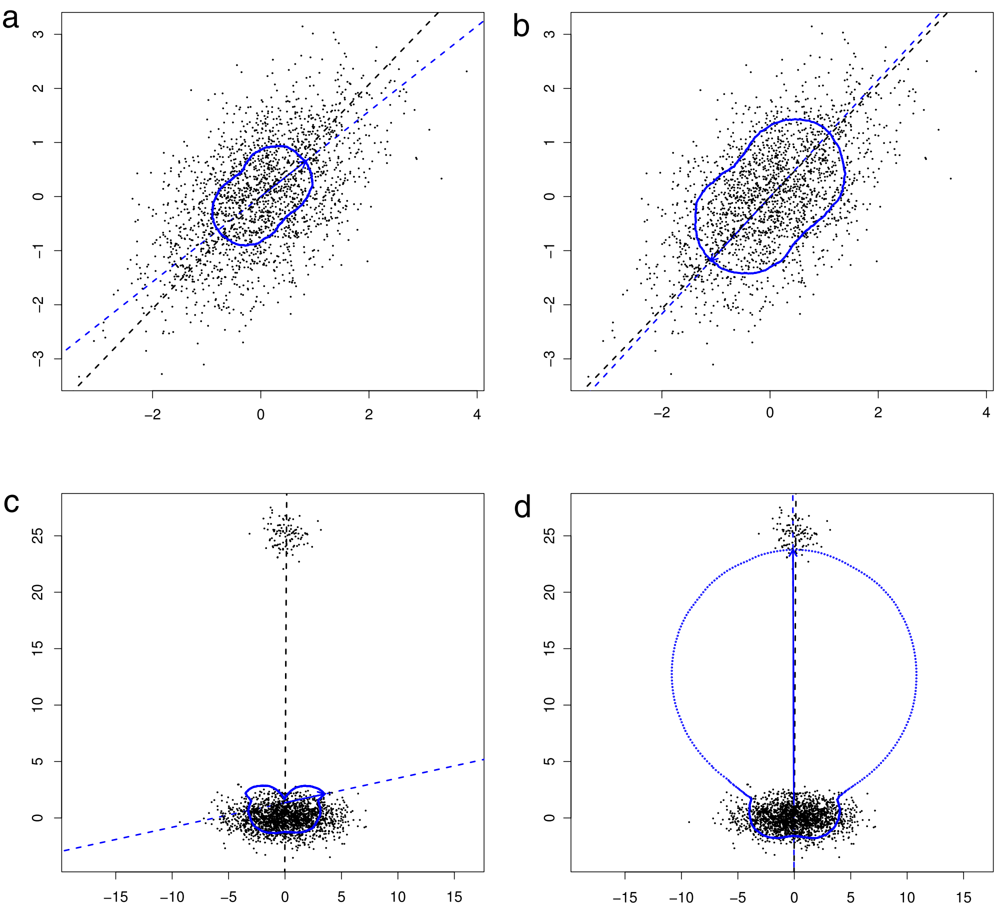

Fig. 2. Top, random sample of size $n = 2000$ from a Normal distribution with zero mean and covariance matrix $\Sigma = (\sigma_{ij})$, $\sigma_{ii} = 1$, $\sigma_{ij} = 0.5$, $i \neq j$. Bottom, random sample of size $n = 2000$ from the mixture of Normal distributions $0.95N(\mu_1, \Sigma_1) + 0.05N(\mu_2, \Sigma_2)$, where $\mu_1 = (0, 0)$, $\Sigma_1 = (\sigma_{ij})$, $\sigma_{11} = 6$, $\sigma_{22} = 1\sigma_{ij} = 0$, $i \neq j$, $\mu_2 = (0, 25)$ and Σ_2 is the identity matrix. In dashed black, line in the direction of the first principal component. In blue, empirical quantile computed over 500 directions and, in dashed blue, line in the direction of the first empirical principal quantile for (a) $\alpha = 0.8$, (b) $\alpha = 0.9$, (c) $\alpha = 0.9$, (d) $\alpha = 0.95$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The second sample principal quantile direction is defined as the direction $\hat{u}_2 \in \mathbb{B}$ satisfying

$$\hat{u}_2 = \operatorname*{arg\,max}_{u \in \mathbb{B}, u \perp \hat{u}_1} |Q(P_n(u), \alpha)|.$$

The sample k-principal quantile direction is defined as the direction $\hat{u}_k \in \mathbb{B}$ satisfying

$$\hat{u}_k = \operatorname*{arg\,max}_{u \in \mathbb{B}, u \perp H_{k-1}} |Q(P_n(u), \alpha)|,$$

where H_{k-1} is the linear subspace generated by $\hat{u}_1, \ldots, \hat{u}_{k-1}$. Fig. 2 show illustrative examples of the behavior of the first principal quantile direction in comparison with the first principal component direction for different samples.

4.1.1. Consistency of principal quantile directions

We will now use the results in Section 4 to prove the strong consistency of the principal quantile directions. Recall that, given $\mathcal{X}$ in $\mathcal{H}$ with probability distribution $P_{\mathcal{X}}$ and $\mathbb{E}(\|\mathcal{X}\|) < \infty$, we denote $\mathcal{Z} = \mathcal{X} - \mathbb{E}(\mathcal{X})$. Now, let

$$\mathbb{F}_1 = \left\{ u \in \mathbb{B} : u = \operatorname*{arg\,max}_{u \in \mathbb{B}} |Q(P_{\mathcal{Z}}(u), \alpha)| \right\},$$

$$\mathbb{F}_{1n} = \left\{ u \in \mathbb{B} : u = \operatorname*{arg\,max}_{u \in \mathbb{B}} |Q(P_n(u), \alpha)| \right\},$$

and consider the following additional assumption.

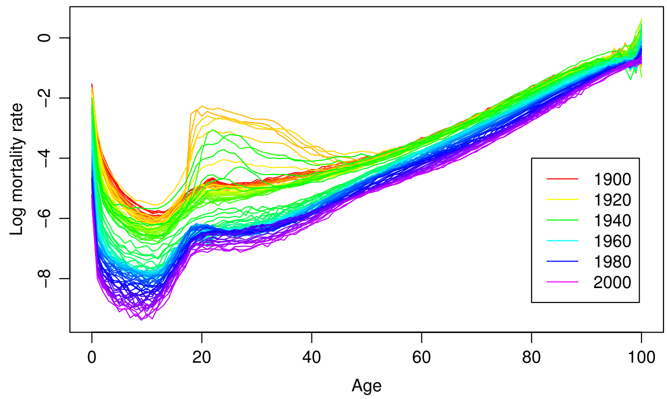

Fig. 3. French male age-specific mortality rates (1899–2005). The color palette is based on the time-ordering of the data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Assumption C1. Given $\epsilon > 0$ and $u_1 \in \mathbb{F}_1$, there exist $\delta > 0$ such that

$$|Q(P_{\mathcal{Z}}(u), \alpha)| < |Q(P_{\mathcal{Z}}(u_1), \alpha)| - \delta, \quad \forall u \notin B(\mathbb{F}_1, \epsilon),$$

where $B(\mathbb{F}_1, \epsilon) = \bigcup_{u \in \mathbb{F}_1} B(u, \epsilon)$, being $B(u, \epsilon)$ the ball with center u and radius ϵ.

Remark 2. In the finite-dimensional case, Assumption C1 will hold if for instance $Q(P_{\mathcal{Z}}(u), \alpha)$ is a continuous function of u. However, this is not the case in the infinite-dimensional setting. The following counterexample shows this fact.

Let $\mathcal{H}$ be a Hilbert space and $\{e_n : n \geq 1\}$ an orthonormal basis of the space, and define $h_1 = e_1$, $h_n = e_n(1 - 1/n)$ for $n \geq 2$. Consider a discrete random element $\mathcal{X}$ verifying $P(\mathcal{X} = h_n) = P(\mathcal{X} = -h_n) = p_n/2$, $\sum_n p_n = 1$. Then we have that

- $\mathbb{E}(\mathcal{X}) = 0$.
- $|\langle \mathcal{Z}, u \rangle| \leq 1$, for $\|u\| = 1$.
- $Q(P_{\mathcal{Z}}(h_1), 1) = 1 = \max_{u \in \mathbb{B}} Q(P_{\mathcal{Z}}(u), 1)$.
- $Q(P_{\mathcal{Z}}(u), 1) < 1$ for $\|u\| = 1$, $u \neq h_1$.
- $Q(P_{\mathcal{Z}}(h_n), 1) = 1 - 1/n, \quad n \geq 2$.
- $\|h_n - h_1\| > 1$,

and therefore Assumption C1 does not hold.

Proposition 7. *Under the conditions in Proposition 2 and Assumption C1 we have that*

(i) *Given $\epsilon > 0$, $u_n \in \mathbb{F}_{1n}$ implies that $u_n \in B(\mathbb{F}_1, \epsilon)$ if $n \geq n_0$ a.s.*
(ii) *If the principal population quantile directions are unique then,*

$$\lim_{n \to \infty} \|\hat{u}_k - u_k\| = 0 \quad a.s.\ \forall k \geq 1.$$

4.2. Some graphical tools based on principal quantile directions. A real data example

Graphical methods for functional data analysis help us to capture important features of the behavior of the data. In recent years particular attention has been paid to the definition of boxplots and outlier detection tools for functional data. See, for instance, [21,9].

Principal quantile directions provide an alternative procedure to visualize data and, in particular, to detect outliers for functional data. Indeed, by looking at the projection of the data on the principal quantile directions, we can identify outliers correctly. To illustrate this reasoning, we consider the annual smoothed age-specific log mortality curves for french males between 1899 and 2005. We refer to [20] for a complete description of the data. The numerical results have been obtained using the R software [31], the rainbow package by Shang and Hyndman [34] and the fda.usc package by Febrero-Bande and O de la Fuente [11]. Fig. 3 represents a rainbow plot where the color palette is based on the time-ordering of the data. As discussed by Hyndman and Shang [21], some of the mortality curves indicate sudden increases in log mortality rates between ages of 20 and 40 for several years. These changes in mortality patterns are consequence of the First and Second World Wars, as well as the Spanish flu epidemic in 1918 and 1919. While a univariate boxplot for the data projected on the first principal component shows no outliers, the univariate boxplot using the data projected on the first principal quantile direction (see Fig. 4 left) identifies as outliers the data corresponding to the years 1914–1919, 1940, 1943–1944. This fact is also clear from the density estimate for the projected data on the first principal quantile direction; see Fig. 4 right. Besides the small bumps associated with the outliers, a bimodal structure is shown in the density plot. Thus, we observe that our method performs well at identifying the outliers, in accordance to the historical information. See [21] for a discussion on outlier detection performances of other methods.

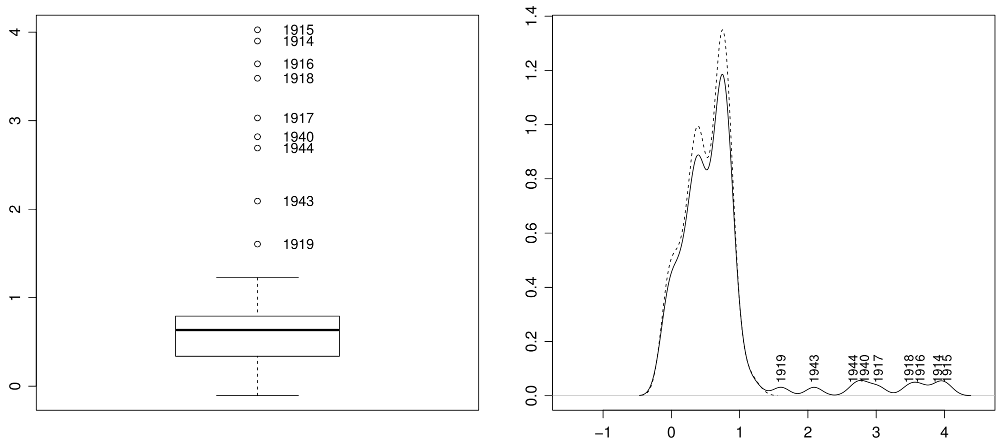

Fig. 4. Left, boxplot of data projected on the first principal quantile direction. The principal quantile direction is computed over 20000 directions with $\alpha = 0.5$. Right, in solid line density estimate for the projected data on the first principal quantile direction. In dashed line, the density estimate constructed without the outliers.

5. Robust principal quantile directions

Robust methods have been developed in the context of multivariate analysis. Also in principal component analysis, there are robust alternatives that emulate the classical methods but are not unduly affected by outliers. The problem of robust principal components for the finite-dimensional case has been considered by several authors starting with [3]. We refer to [29] for discussion on approaches to robust PCA. Robust methods are perhaps more important in functional situations, where the presence of outliers has a serious impact on the modeling. However, for the case of functional data, very few results are available. An important reference in this setting is [26].

The results in the previous section motivate the following simple definition of robust principal components for finite and infinite-dimensional spaces. Given a random element $\mathcal{X} \in \mathcal{H}$, with distribution $P_{\mathcal{X}}$ we consider a robust location functional $T(P_{\mathcal{X}}) =: T_{\mathcal{X}}$ which coincides with μ_0 if the distribution of $\mathcal{X}$ is symmetric about μ_0, like for instance the L^1–median (that minimizes $\mathbb{E}(\|\mathcal{X} - \mu\|)$), or the one defined in [8] (based on the integrated depth over the dual space). For each direction $u \in \mathbb{B}$ consider the interquartile range of the projection on the direction u, defined as

$$\mathrm{ric}_{\mathcal{X}}(\alpha, u) = \tilde{Q}_{\mathcal{X}}\left(\frac{1-\alpha}{2}, u\right) - \tilde{Q}_{\mathcal{X}}\left(\frac{1-\alpha}{2}, -u\right),$$

where

$$\tilde{Q}_{\mathcal{X}}(\alpha, u) = Q_{\langle \mathcal{X} - T_{\mathcal{X}}, u\rangle}(\alpha) u$$

for $\alpha = 0.25$ (for instance), that is a scale measure for the projection of $\mathcal{X}$ on the direction u.

We define robust principal components based on quantiles in the same way as the principal quantile directions obtained when we replace $Q_{\langle \mathcal{X} - \mathbb{E}(\mathcal{X}), u\rangle}(\alpha)$ in (11) with $\mathrm{ric}_{\mathcal{X}}(\alpha, u)$, that is

$$u_1 = \arg\max_{u \in \mathbb{B}} \|\mathrm{ric}_{\mathcal{X}}(\alpha, u)\|.$$

The second principal robust direction is defined as the direction $u_2 \in \mathbb{B}$ satisfying

$$u_2 = \arg\max_{u \in \mathbb{B}, u \perp u_1} \|\mathrm{ric}_{\mathcal{X}}(\alpha, u)\|.$$

The k-principal robust direction is defined as the direction $u_k \in \mathbb{B}$ satisfying

$$u_k = \arg\max_{u \in \mathbb{B}, u \perp H_{k-1}} \|\mathrm{ric}_{\mathcal{X}}(\alpha, u)\|,$$

where H_{k-1} is the linear subspace generated by $u_1, \ldots, u_{k-1}$.

Remark 3. If the distribution of $\mathcal{X}$ is symmetric about $\mathbb{E}(\mathcal{X})$, and $\mathbb{E}(\|\mathcal{X}\|) < \infty$, we have that $\mathcal{X} - \mathbb{E}(\mathcal{X})$ have the same distribution as $\mathbb{E}(\mathcal{X}) - \mathcal{X}$, and the principal robust directions coincide with the principal quantile directions. In particular, the results obtained in the previous section for ellipsoidal and Gaussian distributions still hold.

5.1. Sample robust principal components

Let $T(P_n)$ the robust location estimate obtained when we apply the functional T to the empirical distribution P_n. For instance the empirical L^1-median that minimizes $\sum_{i=1}^{n} \|\mathcal{X}_i - \mu\|$.

For each direction $u \in \mathbb{B}$, let $\tilde{P}_n(u)$ be the empirical distribution of the sample

$$\langle \mathcal{X}_i - T(P_n), u \rangle, \quad i = 1, \ldots, n.$$

Now consider the interquartile range of the projected data on the direction u, defined as

$$\mathrm{ric}(P_n(u), \alpha) = Q\left(\tilde{P}_n(u), \frac{1-\alpha}{2}\right) - Q\left(\tilde{P}_n(-u), \frac{1-\alpha}{2}\right),$$

for $\alpha = 0.25$ (for instance).

The empirical robust principal components based on quantiles are defined now through the equations:

$$\hat{u}_1 = \arg\max_{u\in\mathbb{B}} \|\mathrm{ric}(P_n(u), \alpha)\|.$$

The second robust empirical principal direction is defined as the direction $\hat{u}_2 \in \mathbb{B}$ satisfying

$$\hat{u}_2 = \arg\max_{u\in\mathbb{B}, u\perp\hat{u}_1} \|\mathrm{ric}(P_n(u), \alpha)\|.$$

The robust k-principal direction is defined as the direction $\hat{u}_k \in \mathbb{B}$ satisfying

$$\hat{u}_k = \arg\max_{u\in\mathbb{B}, u\perp H_{n,k-1}} \|\mathrm{ric}(P_n(u), \alpha)\|,$$

where $H_{n,k-1}$ is the linear subspace generated by $\hat{u}_1, \ldots, \hat{u}_{(k-1)}$.

6. Conclusions and discussion

We have introduced a definition of quantile functions for distributions in a Hilbert space. This definition results in quantile contours that enjoy all the desirable properties. They are translation equivariant, scale equivariant and equivariant under unitary transformations (orthogonal equivariant in the finite-dimensional case). In the finite-dimensional case our definition does not lead to affine equivariant quantile contours, but as we have argued, we do not believe this is an adequate property for a multivariate quantile. We have also defined the empirical version of the quantile functions and have shown the uniform strong consistency. We have proposed the empirical version of the principal quantile directions and we have studied their asymptotic properties. Their potential usefulness in the definition of a functional boxplot and in the detection of outliers has been motivated through a real functional data example. Finally, we have seen that a slight modification of the definition of the quantile function can be quite useful to construct robust principal components for functional data.

7. Proofs

Proof of Proposition 1. Given a probability measure P and a function $f : \mathcal{H} \to \mathbb{R}$, denote by Pf the expected value of f under P, that is, $Pf = \int f(x)P(dx)$. Let

$$\mathcal{F} = \{f : \mathcal{H} \to \mathbb{R} : f(x) = \mathbb{I}_{\{\langle h,x\rangle \le t\}},\ \|h\| = 1,\ t \in \mathbb{R}\}.$$

Using the introduced notation, (6) can be written as

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} |P_n f - P_{\mathcal{Z}} f| = 0 \quad \text{a.s.},$$

where P_n denotes the empirical measure of $\mathcal{Z}_{ni}$, $i = 1, \ldots, n$. It is not difficult to prove that the sequence of probability measures P_n converges weakly to the probability measure $P_{\mathcal{Z}}$, that is, for all function f in the class $\mathbb{C}(\mathcal{H})$ of continuous bounded functions in $\mathcal{H}$

$$\lim_{n\to\infty} P_n f = P_{\mathcal{Z}} f \quad \text{a.s.} \tag{14}$$

Let $f \in \mathbb{C}(\mathcal{H})$ and denote by P_n^* the empirical measure of the observations $\mathcal{X}_1 - \mathbb{E}(\mathcal{X}), \ldots, \mathcal{X}_n - \mathbb{E}(\mathcal{X})$. Then,

$$|P_n f - P_{\mathcal{Z}} f| \le \left|P_n f - P_n^* f\right| + \left|P_n^* f - P_{\mathcal{Z}} f\right|. \tag{15}$$

If $P_{\mathcal{Z}}\,|f| < \infty$, then P_n^* converges weakly to $P_{\mathcal{Z}}$. On the other hand, the continuity of f along with the law of large numbers yields that the first term in the right-hand side of (15) converges to zero almost surely and (14) holds. Now, according to the Billingsley–Topsøe Theorem in [2], we can conclude that (6) holds if and only if

$$\lim_{\epsilon\to 0} \sup_{f\in\mathcal{F}} P\left(\{x \in \mathcal{H} : \omega_f(x, \epsilon) > \delta\}\right) = 0 \tag{16}$$

for all $\delta > 0$, where

$$\omega_f(x, \epsilon) = \sup\{|f(x_1) - f(x_2)| : x_1, x_2 \in B(x, \epsilon)\},$$

being $B(x, \epsilon)$ the ball with center x and radius ϵ. Since for $\delta < 1$,

$$\left\{x \in \mathcal{H} : \sup_{x_1, x_2 \in B(x,\epsilon)} \left|\mathbb{I}_{\{\langle h, x_1\rangle \le t\}} - \mathbb{I}_{\{\langle h, x_2\rangle \le t\}}\right| > \delta\right\} = \{x \in \mathcal{H} : |\langle h, x\rangle - t| < \epsilon\},$$

we obtain that (7) is just (16), which concludes the proof. □

Proof of Corollary 1. By Proposition 1, it suffices to prove that (7) holds when $\mathcal{H} = \mathbb{R}^d$. That is, we want to prove that

$$\lim_{\epsilon \to 0} G(\epsilon) = 0,$$

where

$$G(\epsilon) = \sup_{\|\boldsymbol{h}\|=1, t\in\mathbb{R}} P\left(\{\boldsymbol{x} \in \mathbb{R}^d : |\langle \boldsymbol{h}, \boldsymbol{x}\rangle - t| < \epsilon\}\right).$$

Let us first restrict t to a compact set $[-k, k]$, with $k \in \mathbb{R}$. Thus, define

$$G_k(\epsilon) = \sup_{\|\boldsymbol{h}\|=1, |t|\le k} P\left(\{\boldsymbol{x} \in \mathbb{R}^d : |\langle \boldsymbol{h}, \boldsymbol{x}\rangle - t| < \epsilon\}\right). \tag{17}$$

Let $L = \lim_{\epsilon\to 0} G_k(\epsilon)$, which exists since $G_k(\epsilon)$ is a bounded monotone function. Suppose $L > 0$. Then we can define a sequence $\{\epsilon_n\}$ such that $\lim_{n\to\infty} \epsilon_n = 0$ and, for all n,

$$G_k(\epsilon_n) \ge p_0 \tag{18}$$

for some $p_0 > 0$. Since the set $\mathbb{B} \times [-k, k]$ is compact and P is continuous, the supremum in (17) is attained, that is, for each n there exist $(\boldsymbol{h}_0(\epsilon_n), t_0(\epsilon_n))$ in $\mathbb{B} \times [-k, k]$ such that

$$G_k(\epsilon_n) = P\left(\{\boldsymbol{x} \in \mathbb{R}^d : |\langle \boldsymbol{h}_0(\epsilon_n), \boldsymbol{x}\rangle - t_0(\epsilon_n)| < \epsilon_n\}\right).$$

Using again the compactness of $\mathbb{B} \times [-k, k]$, we can define a subsequence $\{\boldsymbol{h}_0(\epsilon_{n_l}), t_0(\epsilon_{n_l})\}$ converging to $(\boldsymbol{h}_0, t_0)$. By (18), $G_k(\epsilon_{n_l}) \ge p_0 > 0$. This yields a contradiction, since for any absolutely continuous probability P,

$$\lim_{l\to\infty} G_k(\epsilon_{n_l}) = \lim_{l\to\infty} P\left(\{\boldsymbol{x} \in \mathbb{R}^d : |\langle \boldsymbol{h}_0, \boldsymbol{x}\rangle - t_0| < \epsilon_{n_l}\}\right) = 0.$$

Therefore, $L = 0$. The proof concludes using that P is tight in $\mathbb{R}^d$. Then, there exists $r > 0$ such that, for any $\delta > 0$, $P(\mathbb{R}^d \setminus B(0, r)) < \delta$. For those sets $\{\boldsymbol{x} \in \mathbb{R}^d : |\langle \boldsymbol{h}, \boldsymbol{x}\rangle - t| < \epsilon\}$ with $|t| > r + 1$ and $\epsilon < 1$ we have

$$P(\{\boldsymbol{x} \in \mathbb{R}^d : |\langle \boldsymbol{h}, \boldsymbol{x}\rangle - t| < \epsilon\}) \le P(\mathbb{R}^d \setminus B(0, r)) < \delta. \quad \square$$

Proof of Proposition 2. Let $0 < \alpha < 1$. Assumption (i), together with the definition of univariate quantile yield that, given $\epsilon > 0$, there exists $\delta_0 > 0$ such that

$$F^u(Q(P_{\mathcal{Z}}(u), \alpha) - \epsilon) < \alpha - \delta_0 \tag{19}$$

and

$$F^u(Q(P_{\mathcal{Z}}(u), \alpha) + \epsilon) > \alpha + \delta_0. \tag{20}$$

Now, by assumption (ii), with probability one and for all $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that, for all $n \ge n_0$

$$|F_n^u(Q(P_n(u), \alpha)) - F^u(Q(P_n(u), \alpha))| < \delta.$$

That is

$$F^u(Q(P_n(u), \alpha)) > \alpha - \delta \tag{21}$$

and

$$F^u(Q(P_n(u), \alpha)) < \alpha + \delta. \tag{22}$$

We obtain that $Q(P_n(u), \alpha) > Q(P_{\mathcal{Z}}(u), \alpha) - \epsilon$, since otherwise Eqs. (19) and (21) would yield a contradiction. Similarly, from (20) and (22) we obtain that $Q(P_n(u), \alpha) < Q(P_{\mathcal{Z}}(u), \alpha) + \epsilon$. This concludes the proof of (8). □

Proof of Proposition 3. By assumption (ii), with probability one and for all $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$

$$\sup_{\|u\|=1} |F_n^u(Q(P_n(u), \alpha)) - F^u(Q(P_n(u), \alpha))| < \delta. \tag{23}$$

From (23) we obtain

$$\sup_{\|u\|=1} F^u(Q(P_n(u), \alpha)) > \alpha - \delta$$

and

$$\inf_{\|u\|=1} F^u(Q(P_n(u), \alpha)) < \alpha + \delta.$$

This together with (9) and (10) concludes the proof by a reasoning similar to that of Proposition 2. $\square$

Proof of Proposition 4. This result is a straightforward consequence of Proposition 3 and the law of large numbers since

$$\begin{aligned} 0 &\leq \sup_{\|u\|=1} \left\| \hat{Q}_{\mathcal{X}}(\alpha, u) - Q_{\mathcal{X}}(u, \alpha) \right\| \\ &\leq \sup_{\|u\|=1} |Q(P_n(u), \alpha) - Q(P_{\mathcal{Z}}(u), \alpha)| + \left\| \bar{\mathcal{X}} - \mathbb{E}(\mathcal{X}) \right\| \end{aligned}$$

and, with probability one, both terms in the right-hand side converge to zero. $\square$

Proof of Proposition 5. By Fact 2, finding the principal quantile directions is equivalent to finding

$$\boldsymbol{u}_k = \underset{\boldsymbol{u}\in\mathbb{B}, \boldsymbol{u}\perp H_{k-1}}{\arg\max}\ \boldsymbol{u}^t \boldsymbol{B}\boldsymbol{B}^t \boldsymbol{u} = \underset{\boldsymbol{u}\in\mathbb{B}, \boldsymbol{u}\perp H_{k-1}}{\arg\max}\ \boldsymbol{u}^t \boldsymbol{\Sigma} \boldsymbol{u},$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the vector $\boldsymbol{X}$. This follows from the fact that the covariance matrix of a spherical random vector is the identity matrix up to some constant multiplier. We obtain, therefore, the equations defining the principal components. $\square$

Proof of Proposition 6. Let $Y(t) = X(t) - \mathbb{E}(X(t))$. For $u \in \mathbb{B}$, denote

$$V_u = \langle \mathcal{X} - \mathbb{E}(\mathcal{X}), u \rangle = \int_0^1 u(t) Y(t) dt.$$

Since $\mathcal{Y} = \{Y(t),\ t \in [0, 1]\}$ is a zero mean Gaussian process, we have that V_u is normally distributed with zero mean and variance σ^2 given by

$$\begin{aligned} \sigma^2 &= \mathbb{E}(V_u^2) \\ &= \mathbb{E}\left(\int_0^1 \int_0^1 u(s)u(t)Y(s)Y(t)dtds \right) \\ &= \int_0^1 \int_0^1 u(s)u(t)\mathbb{E}\left(Y(s)Y(t)\right) dtds \\ &= \int_0^1 \int_0^1 u(s)u(t)\gamma(s,t)dtds = \langle u, \Gamma u \rangle. \end{aligned}$$

Now, since V_u has distribution $N(0, \langle u, \Gamma u \rangle)$, we have that $\left| Q_{\langle \mathcal{X}-\mathbb{E}(\mathcal{X}), u\rangle}(\alpha) \right|$ reaches its maximum in the direction of maximal variance $\langle u, \Gamma u \rangle$. The solution to this problem are the principal components.

Finally, since for the eigenfunctions ϕ_i of the Γ operator we have $\langle \phi_i, \Gamma \phi_i \rangle = \langle \phi_i, \lambda_i \phi_i \rangle = \lambda_i$, then (13) holds. $\square$

Proof of Proposition 7. We prove (i) and (ii).

(i) Proposition 2 entails that

$$\lim_{n\to\infty} |Q(P_n(u), \alpha) - Q(P_{\mathcal{Z}}(u), \alpha)| = 0 \quad \text{a.s.}$$

Now, since $||a| - |b|| \leq |a - b|$, we have that given $u_n \in \mathbb{F}_{1n}$ and if $n \geq n_0(\omega)$,

$$\begin{aligned} |Q(P_{\mathcal{Z}}(u_n), \alpha)| &> |Q(P_n(u_n), \alpha)| - \frac{\delta}{3} \\ &\geq |Q(P_n(u), \alpha)| - \frac{\delta}{3} \end{aligned}$$

for all $u \in \mathbb{F}_1$ (since $u_n \in \mathbb{F}_{1n}$). Then,

$$|Q(P_{\mathbb{Z}}(u_n), \alpha)| \geq |Q(P_n(u), \alpha)| - |Q(P_{\mathbb{Z}}(u), \alpha)| + |Q(P_{\mathbb{Z}}(u), \alpha)| - \frac{\delta}{3}$$
$$\geq Q(P_{\mathbb{Z}}(u), \alpha) - \frac{2\delta}{3},$$

which implies that $u_n \in B(\mathbb{F}_1, \epsilon)$ by Assumption C1.

(ii) From (i) we have that $\lim_{n\to\infty} \|\hat{u}_1 - u_1\| = 0$ a.s. Let $S_1 =< u_1 >^{\perp}$ and $\hat{S}_1 =< \hat{u}_1 >^{\perp}$. As before, by Proposition 2 we have that, for $n \geq n_0(\omega)$,

$$|Q(P_{\mathbb{Z}}(\hat{u}_2), \alpha)| > |Q(P_n(\hat{u}_2), \alpha)| - \frac{\delta}{3}$$
$$= |Q(P_n(\hat{u}_2), \alpha)| - |Q(P_n(u_2), \alpha)| + |Q(P_n(u_2), \alpha)| - \frac{\delta}{3}.$$

If $u_2 \in \hat{S}_1$, then $|Q(P_n(\hat{u}_2), \alpha)| - |Q(P_n(u_2), \alpha)| > 0$ (since the maximum is attained at $\hat{u}_2$) and the proof follows as in (i). Otherwise, $u_2 \notin \hat{S}_1, \ u_2 \in S_1$. But $u_2 = u_{21} + u_{22}$ with $u_{21} \in \hat{S}_1$ and $\lim_{n\to\infty} \|u_{22}\| = 0$ a.s. (by (i)). Then, by the continuity of $|Q(P_n(.), \alpha)|$ we get that

$$|Q(P_{\mathbb{Z}}(\hat{u}_2), \alpha)| > |Q(P_n(u_2), \alpha)| - \frac{2\delta}{3},$$

and again the proof follows as in (i). The proof for $k > 2$ follows exactly in the same way. $\square$

Acknowledgments

The authors gratefully acknowledge the suggestions and comments of the anonymous referees. This research has been partially supported by Spanish grant MTM2010-17366 (first author), Spanish grant MTM2008-03010, and the IAP research network grant no. P6/03 from the Belgian government (second author). The second author also acknowledges the support by an Angeles Alvariño fellowship by Xunta de Galicia.

Appendix. Supplementary data

Supplementary material related to this article can be found online at doi:10.1016/j.jmva.2012.01.016.

References

[1] P.J. Bickel, P.W. Millar, Uniform convergence of probability measures on classes of functions, Statist. Sinica 2 (1) (1992) 1–15.
[2] P. Billingsley, F. Topsøe, Uniformity in weak convergence, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 7 (1967) 1–16.
[3] G. Boente, Asymptotic theory for robust principal components, J. Multivariate Anal. 21 (1) (1987) 67–78.
[4] G. Boente, R. Fraiman, Kernel-based functional principal components, Statist. Probab. Lett. 48 (4) (2000) 335–345.
[5] P. Chaudhuri, On a geometric notion of quantiles for multivariate data, J. Amer. Statist. Assoc. 91 (434) (1996) 862–872.
[6] J.A. Cuesta-Albertos, R. Fraiman, T. Ransford, Random projections and goodness-of-fit tests in infinite-dimensional spaces, Bull. Braz. Math. Soc. (N.S.) 37 (4) (2006) 477–501.
[7] A. Cuevas, M. Febrero, R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions, Comput. Statist. 22 (3) (2007) 481–496.
[8] A. Cuevas, R. Fraiman, On depth measures and dual statistics. A methodology for dealing with general data, J. Multivariate Anal. 100 (4) (2009) 753–766.
[9] M. Febrero, P. Galeano, W. González-Manteiga, A functional analysis of NO_x levels: location and scale estimation and outlier detection, Comput. Statist. 22 (3) (2007) 411–427.
[10] M. Febrero, P. Galeano, W. González-Manteiga, Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels, Environmetrics 19 (4) (2008) 331–345.
[11] M. Febrero-Bande, M. Oviedo de la Fuente, fda.usc: functional data analysis and utilities for statistical computing (fda.usc), r package version 0.9.5 (2011). http://CRAN.R-project.org/package=fda.usc.
[12] F. Ferraty, Statistical methods and problems in infinite-dimensional spaces, J. Multivariate Anal. 101 (2010) Special Issue.
[13] F. Ferraty, Y. Romain, Oxford Handbook on Statistics and Functional Data Analysis, Oxford University Press, Oxford, 2010.
[14] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, in: Springer Series in Statistics, Springer-Verlag New York, Inc., 2006.
[15] R. Fraiman, G. Muniz, Trimmed means for functional data, Test 10 (2) (2001) 419–440.
[16] S. Girard, A nonlinear PCA based on manifold approximation, Comput. Statist. 15 (2) (2000) 145–167.
[17] P. Hall, M. Hosseini-Nasab, On properties of functional principal components analysis, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (1) (2006) 109–126.
[18] P. Hall, H.-G. Müller, J.-L. Wang, Properties of principal component methods for functional and longitudinal data analysis, Ann. Statist. 34 (3) (2006) 1493–1517.
[19] M. Hallin, D. Paindaveine, M. Siman, Multivariate quantiles and multiple-output regression quantiles: from l1 optimization to halfspace depth, Ann. Statist. 38 (2) (2010) 635–669.
[20] R.J. Hyndman, M. Shahid Ullah, Robust forecasting of mortality and fertility rates: a functional data approach, Comput. Statist. Data Anal. 51 (10) (2007) 4942–4956.
[21] R.J. Hyndman, H.L. Shang, Rainbow plots, bagplots and boxplots for functional data, J. Comput. Graph. Statist. 19 (2010) 29–45.
[22] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, 2002.
[23] R. Koenker, G. Bassett Jr., Regression quantiles, Econometrica 46 (1) (1978) 33–50.
[24] L. Kong, I. Mizera, Quantile tomography: Using quantiles with multivariate data, Statistica Sinica, doi:10.5705/ss.2010.224.
[25] R.Y. Liu, J.M. Parelius, K. Singh, Multivariate analysis by data depth: descriptive statistics, graphics and inference, Ann. Statist. 27 (3) (1999) 783–858.

[26] N. Locantore, J.S. Marron, D.G. Simpson, N. Tripoli, J.T. Zhang, K.L. Cohen, Robust principal component analysis for functional data, TEST 8 (1) (1999) 1–73.
[27] S. López-Pintado, J. Romo, Depth-based inference for functional data, Comput. Statist. Data Anal. 51 (10) (2007) 4957–4968.
[28] S. López-Pintado, J. Romo, On the concept of depth for functional data, J. Amer. Statist. Assoc. 104 (486) (2009) 718–734.
[29] R.A. Maronna, R.D. Martin, V.J. Yohai, Robust Statistics, John Wiley & Sons Ltd, 2006.
[30] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, 2005.
[31] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, 2010, http://www.R-project.org/.
[32] R. Serfling, Quantile functions for multivariate analysis: approaches and applications, Stat. Neerl. 56 (2) (2002) 214–232.
[33] R.J. Serfling, Multivariate symmetry and asymmetry, in: C.B.R. S. Kotz, N. Balakrishnan, B. Vidakovic (Eds.), Encyclopedia of Statistical Sciences, Vol. 8, Second Edition, Wiley, 2006.
[34] H.L. Shang, R.J. Hyndman, Rainbow: rainbow plots, bagplots and boxplots for functional data, r package version 2.3.4 (2011). http://CRAN.R-project.org/package=rainbow.
[35] Y. Zuo, R. Serfling, General notions of statistical depth function, Ann. Statist. 28 (2) (2000) 461–482.